Biometrika (2000), **87**, 1, *pp.* 73–87

Printed in Great Britain

Additive hazards regression for case-cohort studies

By MICHAL KULICH

Department of Probability and Statistics, Charles University, Sokolovská 83, 186 75 Prague, Czech Republic

kulich@karlin.mff.cuni.cz

AND D. Y. LIN

Department of Biostatistics, Box 357232, University of Washington, Seattle, Washington 98195, U.S.A.

danyu@biostat.washington.edu

Summary

The case-cohort design is a common means of reducing cost in large epidemiological cohort studies. Under this design, covariates are measured only on the cases and a subcohort randomly selected from the entire cohort. In this paper, we demonstrate how to use the case-cohort data to estimate the regression parameter of the additive hazards model, which specifies that the conditional hazard function given a set of covariates is the sum of an arbitrary baseline hazard function and a regression function of the covariates. The proposed estimator is shown to be consistent and asymptotically normal with an easily estimated variance. The subcohort may be selected by independent Bernoulli sampling with arbitrary selection probabilities or by stratified simple random sampling. The efficiencies of various sampling schemes are investigated both analytically and by simulation. A real example is provided.

Some key words: Censoring; Proportional hazards; Pseudo-score; Risk difference; Stratified sampling; Survival data.

1. Introduction

Cohort studies and prevention trials typically involve the follow-up of several thousand subjects for many years. The assembly of covariate histories can be prohibitively expensive if it is done on all cohort members. Under the case-cohort design (Prentice, 1986), covariate histories are ascertained only for the cases, i.e. those who experience the disease of interest during the follow-up period, and for a small subcohort, which is a random sample from the entire cohort. The reduction of cost offered by this design has enabled researchers to conduct studies that otherwise would have been infeasible.

Prentice (1986) and Self & Prentice (1988) discussed how to use case-cohort data to estimate the relative risk of the proportional hazards model. Their estimators were obtained by approximating the risk sets of the entire cohort involved in the partial likelihood function with their subcohort counterparts. They demonstrated that the efficiency losses of the resulting estimators relative to the maximum partial likelihood estimator based on full covariate data are minimal, particularly for large cohorts with infrequent disease occurrence. Kalbfleisch & Lawless (1988), among others, have suggested that the subcohort may be selected from the entire cohort with unequal probabilities and also that the efficiency of the estimation may be improved by including all the cases in the approximate risk sets even if they do not belong to the subcohort. The properties of these modified designs and estimators have not yet been studied.

All the aforementioned work deals with the proportional hazards regression, which pertains to the relative risk of the exposure. Epidemiologists are also interested in the risk difference attributed to the exposure. The risk difference is more relevant to public health because it translates directly into the number of disease cases that would be avoided by eliminating a particular exposure. The analogue of the proportional hazards model for the risk difference regression is the additive hazards model, which specifies that the hazard function associated with a set of possibly time-dependent covariates $Z(.)$ is given by

$$\lambda(t \mid Z) = \lambda_0(t) + \beta_0^{\mathrm{T}} Z(t), \tag{1}$$

where λ_0 is an unspecified baseline hazard function and β_0 is a vector-valued regression parameter (Cox & Oakes, 1984, p. 74; Breslow & Day, 1987, p. 182).

Lin & Ying (1994) proposed an estimator for β_0 of model (1). Their estimator, however, requires that the covariate data be fully observed. Furthermore, their estimator differs from the maximum partial likelihood estimator of the proportional hazards model in how the covariates of the controls, i.e. disease-free subjects, enter into the estimating function. Therefore, it is not obvious how to estimate β_0 of model (1) based on case-cohort data or whether or not the resulting estimator will have relative efficiency similar to that of the Prentice and Self–Prentice estimators.

This paper provides the answers to the above two questions. Specifically, we discuss in §§ 2 and 3 how to construct appropriate estimators for model (1) when the subcohort is selected by independent Bernoulli sampling with arbitrary selection probabilities or by stratified simple random sampling with fixed sample size. In both settings, the proposed estimators are proven to be consistent and asymptotically normal with easily estimated limiting covariance matrices. In § 4, the asymptotic relative efficiency of the proposed estimator is calculated under various sampling schemes and is compared with that of the Prentice and Self–Prentice estimators. In § 5, the results of a simulation study are reported. A real example is given in § 6.

2. CASE-COHORT DESIGN WITH BERNOULLI SAMPLING

2·1. *Estimation of* β_0

Let T be the failure time, C be the censoring time and $Z(t)$ $(0 \leqslant t \leqslant \tau)$ be a vector of covariate processes, where $\tau < \infty$ denotes the time when the follow-up ends. Write $X = \min(T, C)$ and $\delta = I(T \leqslant C)$, where $I(\mathscr{A})$ is the indicator function of event $\mathscr{A}$. Suppose that the cohort consists of n independent subjects such that $\{T_i, C_i, Z_i(.)\}$ $(i = 1, \ldots, n)$ are n independent copies of $\{T, C, Z(.)\}$.

If covariates were measured on the entire cohort, the data would be

$$\{X_i, \delta_i, Z_i(t), 0 \leqslant t \leqslant X_i\} \quad (i = 1, \ldots, n).$$

Under the case-cohort design, however, covariates are available only on the cases, i.e. those with $\delta_i = 1$, and on a random subset of the entire cohort, i.e. the subcohort; the two sets may overlap. Let ξ_i indicate, by the values 1 versus 0, whether or not the ith subject in the original cohort is selected into the subcohort. In this section, the ξ_i's are independent

Bernoulli variables with possibly unequal success probabilities. We allow ξ_i to depend on individual characteristics V_i, which may involve X_i, $Z_i(.)$ and some external variables correlated with X_i and $Z_i(.)$. Let $p_i = \text{pr}(\xi_i = 1) = p(V_i)$, where $p(V_i)$ is a function mapping the sample space of V to $(p_0, 1)$ for some $p_0 > 0$. The size of the subcohort $\tilde{n} := \sum \xi_i$ is random. However, if $n^{-1} \sum p_i$ converges to $\alpha \in (0, 1]$ in probability, then $\tilde{n}/n$ also converges to α in probability and α is the limiting subcohort proportion. The observable data for the ith subject is $\{X_i, \delta_i, V_i, \xi_i, Z_i(t), 0 \leqslant t \leqslant X_i\}$ if $\xi_i = 1$ or $\delta_i = 1$, and is (δ_i, V_i, ξ_i) if $\xi_i = \delta_i = 0$.

We adopt the standard counting-process notation: $N_i(t) = I(T_i \leqslant t, \delta_i = 1)$ and $Y_i(t) = I(X_i \geqslant t)$. If the data were completely observed, then β_0 of model (1) could be estimated by $\hat{\beta}_A$, the root of the pseudo-score function

$$U_A(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \{Z_i(t) - \bar{Z}(t)\}\{dN_i(t) - \beta^{\mathrm{T}} Z_i(t) Y_i(t)\, dt\},$$

where

$$\bar{Z}(t) = \sum_{i=1}^{n} Z_i(t) Y_i(t) \Big/ \sum_{i=1}^{n} Y_i(t)$$

(Lin & Ying, 1994). A fundamental difference between U_A and the partial likelihood score function of the proportional hazards model is that the former includes one term for each subject no matter if he/she is a case or not, whereas the latter includes the cases only. Thus, an approach different from that of Prentice (1986) and Self & Prentice (1988) is required to adapt U_A to the case-cohort design.

To obtain a pseudo-score for fitting model (1) to case-cohort data, we define the weighted availability indicators $\rho_i = \delta_i + (1 - \delta_i)\xi_i / p_i$ $(i = 1, \ldots, n)$. Note that ρ_i weights the ith subject by the inverse probability of selection, which is set to 1 for all the cases, and that $E(\rho_i \mid \delta_i, V_i) = 1$. Mimicking the Horvitz & Thompson (1951) idea, we propose to modify U_A as

$$U_H(\beta) := \sum_{i=1}^{n} \rho_i \int_0^{\tau} \{Z_i(t) - \bar{Z}_H(t)\}\{dN_i(t) - \beta^{\mathrm{T}} Z_i(t) Y_i(t)\, dt\}, \tag{2}$$

where

$$\bar{Z}_H(t) = \sum_{i=1}^{n} \rho_i Z_i(t) Y_i(t) \Big/ \sum_{i=1}^{n} \rho_i Y_i(t).$$

The resulting estimator possesses a closed form:

$$\hat{\beta}_H = \left[\sum_{i=1}^{n} \rho_i \int_0^{\tau} \{Z_i(t) - \bar{Z}_H(t)\}^{\otimes 2} Y_i(t)\, dt \right]^{-1} \sum_{i=1}^{n} \int_0^{\tau} \{Z_i(t) - \bar{Z}_H(t)\}\, dN_i(t), \tag{3}$$

where $a^{\otimes 2} = aa^{\mathrm{T}}$.

As compared with the Prentice and Self–Prentice pseudo-scores for proportional hazards regression, U_H is based on a much smaller number of terms than the original pseudo-score U_A and thus may incur greater efficiency loss. Unlike its counterpart in the Prentice and Self–Prentice pseudo-scores, $\bar{Z}_H$ includes not only the subcohort members but also all the cases not belonging to the subcohort. This approach shares the same spirit as that of Kalbfleisch & Lawless (1988).

2·2. *Asymptotic distribution of $\hat{\beta}_H$*

Define

$$\Lambda_0(t) = \int_0^t \lambda_0(s)\, ds,$$

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)\, d\Lambda_0(s) - \int_0^t \beta_0^{\mathrm{T}} Z_i(s) Y_i(s)\, ds.$$

Simple algebraic manipulation yields

$$U_H(\beta_0) = \sum_{i=1}^{n} \rho_i \int_0^\tau \{Z_i(t) - \bar{Z}_H(t)\}\, dM_i(t). \tag{4}$$

Although $U_H(\beta_0)$ is the sum of stochastic integrals with respect to martingales, the familiar martingale central limit theorem (Andersen & Gill, 1982) cannot be applied to $U_H(\beta_0)$ because the ρ_i's depend on the δ_i's and are therefore not predictable. We show in Appendix 1 that

$$n^{-\frac{1}{2}} U_H(\beta_0) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \rho_i S_i(\beta_0) + o_p(1),$$

where

$$S_i(\beta_0) = \int_0^\tau \{Z_i(t) - e(t)\}\, dM_i(t),$$

and $e(t) = E\{Z_1(t)Y_1(t)\}/E\{Y_1(t)\}$. The proof involves techniques not previously used in the literature of case-cohort designs.

Since $1 - \rho_i = (1 - \delta_i)(1 - \xi_i/p_i)$, we have

$$n^{-\frac{1}{2}} U_H(\beta_0) = n^{-\frac{1}{2}} \sum_{i=1}^{n} S_i(\beta_0) - n^{-\frac{1}{2}} \sum_{i=1}^{n} (1 - \delta_i)(1 - \xi_i/p_i) S_i(\beta_0) + o_p(1). \tag{5}$$

The two terms on the right-hand side of (5) are both sums of n independent zero-mean random vectors. The first term is asymptotically equivalent to the full-cohort pseudo-score $n^{-\frac{1}{2}} U_A(\beta_0)$, which converges in distribution to a zero-mean normal random vector with covariance matrix

$$\Sigma_A(\beta_0) = E\left[\int_0^\tau \{Z_1(t) - e(t)\}^{\otimes 2}\, dN_1(t)\right].$$

The second term has zero mean conditional on $\{X_i, \delta_i, Z_i(.), V_i\}$ $(i = 1, \ldots, n)$ and converges in distribution to a zero-mean normal random vector with covariance matrix

$$\Sigma_H(\beta_0) = E\{(1 - p_1)p_1^{-1}(1 - \delta_1) S_1^{\otimes 2}(\beta_0)\}.$$

The two terms are clearly uncorrelated. Thus, $n^{-\frac{1}{2}} U_H(\beta_0)$ converges in distribution to a zero-mean normal random vector with covariance matrix $\Sigma_A(\beta_0) + \Sigma_H(\beta_0)$. Loosely speaking, the case-cohort design adds some extra variability Σ_H to the full-cohort pseudo-score covariance matrix Σ_A. Self & Prentice (1988) observed a similar phenomenon for proportional hazards regression.

By Taylor series expansion, $n^{\frac{1}{2}}(\hat{\beta}_H - \beta_0)$ converges in distribution to a zero-mean normal random vector with covariance matrix $D_A^{-1}(\Sigma_A+\Sigma_H)D_A^{-1}$, where

$$D_A=E\left[\int_0^\tau \{Z_1(t)-e(t)\}^{\otimes 2}Y_1(t)\,dt\right],$$

which is the probability limit of $-n^{-1}\,\partial U_H(\beta_0)/\partial\beta$. Since the pseudo-score is linear in β, the consistency of $\hat\beta_H$ follows from the asymptotic normality.

2·3. *Estimation of* $\Lambda_0(.)$ *and variance estimation for* $\hat\beta_H$

A natural estimator for the cumulative baseline hazard $\Lambda_0(t)$ is

$$\hat\Lambda_H(t):=\int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\sum_{j=1}^n \rho_j Y_j(s)}-\int_0^t \hat\beta_H^{\mathrm{T}}\bar Z_H(s)\,ds,$$

which is a simple modification of the estimator proposed by Lin & Ying (1994). It is shown in Appendix 2 that $n^{\frac{1}{2}}\{\hat\Lambda_H(t)-\Lambda_0(t)\}$ converges weakly on $[0,\tau]$ to a zero-mean Gaussian process whose covariance function at (s,t) is

$$h^{\mathrm{T}}(s)D_A^{-1}(\Sigma_A+\Sigma_H)D_A^{-1}h(t)+R_1(s,t)-h^{\mathrm{T}}(s)D_A^{-1}R_2(t)-h^{\mathrm{T}}(t)D_A^{-1}R_2(s),$$

where

$$h(t)=\int_0^t e(u)\,du,$$

$$R_1(s,t)=E\left[\{\delta_1+(1-\delta_1)/p_1\}\int_0^s \pi_0^{-1}(u)\,dM_1(u)\int_0^t \pi_0^{-1}(v)\,dM_1(v)\right],$$

$$R_2(t)=E\left[\int_0^t \{Z_1(u)-e(u)\}\pi_0^{-1}(u)\,dN_1(u)\right],$$

and $\pi_0(t)=\mathrm{pr}(X_1\geqslant t)$.

The full-cohort covariance matrix $\Sigma_A(\beta_0)$ can be consistently estimated by

$$\hat\Sigma_A:=n^{-1}\sum_{i=1}^n\int_0^\tau \{Z_i(t)-\bar Z_H(t)\}^{\otimes 2}\,dN_i(t). \tag{6}$$

A consistent estimator for D_A is

$$\hat D_A:=n^{-1}\sum_{i=1}^n \rho_i\int_0^\tau \{Z_i(t)-\bar Z_H(t)\}^{\otimes 2}Y_i(t)\,dt. \tag{7}$$

The consistency of $\hat\Sigma_A$ and $\hat D_A$ follows from the law of large numbers, together with the uniform convergence of $\bar Z_H(t)$ to $e(t)$ established in Appendix 1.

To estimate the extra pseudo-score covariance Σ_H, we define

$$\hat\Sigma_H(\hat\beta_H)=n^{-1}\sum_{i=1}^n \frac{1-p_i}{p_i^2}\,\xi_i(1-\delta_i)\hat S_i^{\otimes 2}(\hat\beta_H),$$

where

$$\hat S_i(\hat\beta_H)=\int_0^\tau \{Z_i(t)-\bar Z_H(t)\}\{dN_i(t)-Y_i(t)\,d\hat\Lambda_H(t)-\hat\beta_H^{\mathrm{T}}Z_i(t)Y_i(t)\,dt\},$$

which is obtained from $S_i(\beta_0)$ by replacing all the unknown parameters with their respective sample estimators. Note that $\hat{\Sigma}_H(\hat{\beta}_H)$ involves only those $\hat{S}_i$'s associated with the controls, i.e. zero δ_i's. When $\delta_i = 0$,

$$\hat{S}_i(\hat{\beta}_H) = -\int_0^\tau \{Z_i(t) - \bar{Z}_H(t)\} \frac{Y_i(t)}{\sum_{j=1}^n \rho_j Y_j(t)} \left\{ \sum_{k=1}^n dN_k(t) \right\} - \int_0^\tau \{Z_i(t) - \bar{Z}_H(t)\}^{\otimes 2} \hat{\beta}_H Y_i(t)\, dt.$$

The proof of the consistency of $\hat{\Sigma}_H(\hat{\beta}_H)$ is technical, but follows from the uniform convergence of $\bar{Z}_H(t)$ and $\hat{\Lambda}_H(t)$ and the repeated use of integration by parts.

3. Case-cohort design with simple random sampling

The sampling scheme considered in the previous section is very flexible. If the selection probabilities are equal for all subjects or for subjects within the same stratum, however, efficiency can be improved by using simple random sampling with fixed sample size. Thus, we consider in this section possibly stratified simple random sampling for the case-cohort design.

Suppose that the cohort is divided into $K \geqslant 1$ strata of sizes $n_1, \ldots, n_K$ according to certain criteria, where $n_1 + \ldots + n_K = n$. Let $1 \leqslant i \leqslant n_k$ index the subjects within the kth stratum. From the n_k subjects in the kth stratum, we select at random $\tilde{n}_k$ subjects into the subcohort. The total subcohort size is $\tilde{n} := \tilde{n}_1 + \ldots + \tilde{n}_K$. Denote by $p_k := \tilde{n}_k / n_k$ the proportion sampled from the kth stratum and assume that p_k converges to $\alpha_k \in (0, 1]$ as $n \to \infty$. Write q_k for the limit of n_k/n. The total subcohort proportion $\tilde{n}/n$ then converges to $\alpha := \sum q_k \alpha_k$.

Let ξ_{ki} indicate whether or not the ith subject of the kth stratum is selected into the subcohort. Unlike in the case of Bernoulli sampling $\xi_{k1}, \ldots, \xi_{kn_k}$ are correlated because of the fixed sample size $\tilde{n}_k$. Define $\rho_{ki} = \delta_{ki} + (1 - \delta_{ki})\xi_{ki}/p_k$. A natural analogue of pseudo-score (2) is

$$U_H(\beta) := \sum_{k=1}^K \sum_{i=1}^{n_k} \rho_{ki} \int_0^\tau \{Z_{ki}(t) - \bar{Z}_H(t)\} \{dN_{ki}(t) - \beta^{\mathrm{T}} Z_{ki}(t) Y_{ki}(t)\, dt\}. \qquad (8)$$

The resulting estimator takes a similar form to (3).

To establish the asymptotic distribution of (8), we use an approximation similar to (5). Specifically,

$$n^{-\frac{1}{2}} U_H(\beta_0) = n^{-\frac{1}{2}} \sum_{k=1}^K \sum_{j=1}^{n_k} S_{kj}(\beta_0) - n^{-\frac{1}{2}} \sum_{k=1}^K \sum_{j=1}^{n_k} (1 - \delta_{kj}) \left(1 - \frac{\xi_{kj}}{p_k}\right) S_{kj}(\beta_0) + o_P(1), \quad (9)$$

where

$$S_{ki}(\beta_0) = \int_0^\tau \{Z_{ki}(t) - e(t)\}\, dM_{ki}(t).$$

The proof of (9) is even more delicate than that of (5) because of the dependence among the ξ_{kj} $(j = 1, \ldots, n_k)$, but is given in Appendix 1.

The first term on the right-hand side of (9) is identical to that of (5). By a slight extension of Hájek's (1960) central limit theorem for simple random sampling, the second term converges in distribution to a zero-mean normal random vector with covariance matrix

$$\Sigma_H^*(\beta_0)=\sum_{k=1}^{K} q_k \frac{1-\alpha_k}{\alpha_k}\Sigma_k^*(\beta_0),$$

where

$$\Sigma_k^*(\beta_0)=E_k\{(1-\delta_{ki})S_{ki}^{\otimes 2}(\beta_0)\}-[E_k\{(1-\delta_{ki})S_{ki}(\beta_0)\}]^{\otimes 2},$$

and E_k denotes expectation taken over the kth stratum. Therefore, $n^{-\frac{1}{2}}U_H(\beta_0)$ is asymptotically zero-mean normal with covariance matrix $\Sigma_A(\beta_0)+\Sigma_H^*(\beta_0)$ and consequently $n^{-\frac{1}{2}}(\hat{\beta}_H-\beta_0)$ is asymptotically zero-mean normal with covariance matrix $D_A^{-1}(\Sigma_A+\Sigma_H^*)D_A^{-1}$. Since in general $E_k\{(1-\delta_{ki})S_{ki}(\beta_0)\}\neq 0$, the extra covariance $\Sigma_H^*(\beta_0)$ under stratified simple random sampling is never larger than the extra covariance $\Sigma_H(\beta_0)$ under stratified Bernoulli sampling. This holds for all $K\geqslant 1$.

It is natural to estimate $\Sigma_k^*(\beta_0)$ by

$$\hat{\Sigma}_k^*(\hat{\beta}_H)=(n_k p_k)^{-1}\sum_{i=1}^{n_k}\xi_{ki}(1-\delta_{ki})\hat{S}_{ki}^{\otimes 2}(\hat{\beta}_H)-\left\{(n_k p_k)^{-1}\sum_{i=1}^{n_k}\xi_{ki}(1-\delta_{ki})\hat{S}_{ki}(\hat{\beta}_H)\right\}^{\otimes 2},$$

where

$$\hat{S}_{ki}(\hat{\beta}_H)=-\int_0^{\tau}\{Z_{ki}(t)-\bar{Z}_H(t)\}\frac{Y_{ki}(t)}{\sum_{l,m}Y_{lm}(t)}\,d\sum_{l,m}N_{lm}(t)$$
$$-\int_0^{\tau}\{Z_{ki}(t)-\bar{Z}_H(t)\}^{\otimes 2}Y_{ki}(t)\hat{\beta}_H\,dt.$$

Thus, the limiting covariance matrix of $n^{-\frac{1}{2}}(\hat{\beta}_H-\beta_0)$ can be estimated by $\hat{D}_A^{-1}(\hat{\Sigma}_A+\hat{\Sigma}_H^*)\hat{D}_A^{-1}$, where

$$\hat{\Sigma}_H^*(\hat{\beta}_H)=\sum_{k=1}^{K}\frac{n_k(1-p_k)}{np_k}\hat{\Sigma}_k^*(\hat{\beta}_H),$$

and $\hat{\Sigma}_A$ and $\hat{D}_A$ are obvious modifications of (6) and (7), respectively.

4. Asymptotic relative efficiency

The goal of this section is to address three related issues. First, what is the efficiency loss of the case-cohort estimator $\hat{\beta}_H$ relative to the full-data estimator $\hat{\beta}_A$ and how does it compare to the efficiency loss of the Self–Prentice estimator? Secondly, what is the efficiency gain of stratified simple random sampling over stratified Bernoulli sampling? Finally, by how much can the efficiency be improved by stratifying on a surrogate exposure?

Let Z be a binary time-independent exposure with $\mathrm{pr}(Z=1)=p_Z$. We consider stratified subcohort sampling from two strata defined by a dichotomous surrogate exposure V. Write

$$\eta=\mathrm{pr}(V=1\,|\,Z=1),\quad \nu=\mathrm{pr}(V=0\,|\,Z=0),\quad p_V:=\mathrm{pr}(V=1)=(1-\nu)(1-p_Z)+\eta p_Z,$$

where η is the sensitivity and ν is the specificity of the surrogate for the true exposure. Assume that V is independent of all other variables given Z. The goal of the stratification is to achieve a more balanced subcohort with respect to Z by using the information on V.

How can this be done? For stratified simple random sampling, we just select $\tilde{n}_0=[n\alpha/2]$ subjects from each stratum. Denote the resulting extra pseudo-score variance by Σ_{HS}. To conduct stratified Bernoulli sampling, set $p(V)=\alpha/(2p_V)$ for $V=1$ and $p(V)=\alpha/(2-2p_V)$ for $V=0$. The extra pseudo-score variance is denoted by Σ_{HB}. Note that an unstratified sample with equal probabilities can be obtained as a special case if we set $\eta=\nu=\frac{1}{2}$.

In Appendix 3, we present formulae for Σ_A, Σ_{HS} and Σ_{HB} assuming that all the censoring occurs at $\tau=1$. As shown in § 3, Σ_{HS} is never smaller than Σ_{HB}. It follows from Appendix 3 that, in the current setting, $\Sigma_{HS}=\Sigma_{HB}$ if and only if both $\beta_0=0$ and $\eta+\nu=1$. This implies that stratified random sampling yields more efficient estimators whenever $\beta_0\neq 0$ or V is correlated with Z.

We used the formulae given in Appendix 3 to evaluate the asymptotic relative efficiencies of various pairs of estimators in the special case of constant baseline hazard. Since the proportional hazards model also holds, we evaluated the efficiency of the Self–Prentice estimator as well; the Prentice estimator is asymptotically equivalent to the Self–Prentice estimator. This calculation was based on equations (5·7)–(5·9) from Self & Prentice (1988). The results are shown in Table 1. Note that, for $\beta_0=0$, the results apply to any baseline hazard function because the asymptotic variances then depend only on $\Lambda_0(1)$.

Table 1(a) compares the efficiencies of unstratified $\hat{\beta}_H$ under Bernoulli sampling and

Table 1. *Asymptotic relative efficiencies under unstratified and stratified designs*

(*a*) *Unstratified versus full*, $p_Z=0{\cdot}3$

C/C	Rel. risk	ARE$_1$ Bern.	ARE$_2$ s.r.s.	ARE$_3$ S–P	C/C	Rel. risk	ARE$_1$ Bern.	ARE$_2$ s.r.s.	ARE$_3$ S–P
		$E(\delta_1)=0{\cdot}01$					$E(\delta_1)=0{\cdot}10$		
1	1	0·503	0·503	0·502	1	1	0·527	0·527	0·518
	3	0·371	0·402	0·461		3	0·398	0·423	0·472
2	1	0·671	0·671	0·670	2	1	0·714	0·714	0·707
	3	0·543	0·576	0·634		3	0·598	0·623	0·668
4	1	0·806	0·806	0·806	4	1	0·870	0·870	0·866
	3	0·708	0·735	0·779		3	0·799	0·815	0·843

(*b*) *Stratified versus unstratified*, $E(\delta_i)=0{\cdot}01$

C/C	Rel. risk	$\eta=\nu$	ARE$_4$ Bern.	ARE$_5$ s.r.s.	C/C	Rel. risk	$\eta=\nu$	ARE$_4$ Bern.	ARE$_5$ s.r.s.
		$p_Z=0{\cdot}1$					$p_Z=0{\cdot}3$		
1	1	0·7	1·054	1·087	1	1	0·7	1·013	1·086
		0·9	1·259	1·467			0·9	1·054	1·462
	3	0·7	1·098	1·126		3	0·7	1·038	1·104
		0·9	1·553	1·803			0·9	1·174	1·608
2	1	0·7	1·036	1·056	2	1	0·7	1·009	1·055
		0·9	1·159	1·268			0·9	1·036	1·264
	3	0·7	1·074	1·094		3	0·7	1·028	1·072
		0·9	1·383	1·519			0·9	1·122	1·367

ARE$_1$ and ARE$_2$ are the asymptotic efficiencies of $\hat{\beta}_H$ relative to the full-data estimator under Bernoulli sampling and simple random sampling, respectively; ARE$_3$ is the asymptotic efficiency of the Self–Prentice estimator relative to the full-data estimator; ARE$_4$ and ARE$_5$ are the asymptotic relative efficiencies of stratified versus unstratified designs with Bernoulli sampling and simple random sampling, respectively. C/C, number of controls per case.

simple random sampling with the Self–Prentice estimator, which has the same asymptotic variance under both types of sampling. As indicated above, simple random sampling is more efficient than Bernoulli sampling if the exposure affects the failure time. Both estimators are as efficient as the Self–Prentice estimator under the null; the small discrepancy is caused by the exclusion of the cases not belonging to the subcohort from the approximate risk sets by the Self–Prentice estimator. As suspected, the Self–Prentice estimator becomes more efficient as the exposure effect increases, but the difference is not very large even with the relative risk of 3.

Table 1(b) also evaluates the effect of stratification. With either Bernoulli or simple random sampling, the efficiency gain due to stratification is substantial when the surrogate is good, the exposure is rare and its effect is large. However, with Bernoulli sampling the gain tapers off much faster as the exposure gets more common.

5. Simulation study

To investigate the behaviour of the proposed estimators for sample sizes commonly encountered in practice, we conducted a Monte Carlo experiment. The set-up and notation are the same as in § 4 unless otherwise indicated. We generated samples of 5000 subjects with failure times satisfying the model $\lambda(t\,|\,Z) = 0{\cdot}5 + \beta_0 Z$, where Z is binary with $p_Z = 0{\cdot}1$ or $0{\cdot}5$. Unlike in § 4, the censoring was uniform over a certain time interval. The proportion of cases was close to 0·1 and the subcohort proportion was 0·1. The subcohort was selected by simple random sampling, with or without stratification. In each case, 1000 simulation samples were generated.

Table 2 displays the Monte Carlo estimates for the sampling means and sampling standard errors of the estimators, for the sampling means of the standard error estimators, for the coverage percentages of 95% confidence intervals and for the relative efficiencies. The results for efficiencies confirm the conclusion reached in § 4: stratification always pays off, especially when the surrogate is precise. The sampling standard errors of the parameter estimates are close to the average standard error estimates and the corresponding 95%

Table 2. *Simulation results for simple random sampling of the subcohort under model* $\lambda(t\,|\,Z) = 0{\cdot}5 + \beta_0 Z$

	$p_Z = 0{\cdot}1$					$p_Z = 0{\cdot}5$				
Est.	Mean of est.	SE of est.	Mean SEE	Cover. (%)	RE	Mean of est.	SE of est.	Mean SEE	Cover. (%)	RE
					$\beta_0 = 0$					
F	0·002	0·078	0·076	94·4	1·00	−0·001	0·046	0·045	95·6	1·00
CC	0·014	0·112	0·117	96·8	0·42	0·002	0·068	0·067	95·2	0·45
S_1	0·011	0·113	0·110	93·9	0·47	0·002	0·064	0·065	96·2	0·48
S_2	0·005	0·095	0·092	93·5	0·67	0·002	0·058	0·059	95·3	0·59
					$\beta_0 = 0{\cdot}5$					
F	0·500	0·117	0·111	93·8	1·00	0·498	0·068	0·068	94·9	1·00
CC	0·531	0·201	0·201	95·6	0·31	0·508	0·106	0·102	94·9	0·44
S_1	0·513	0·186	0·184	93·4	0·37	0·504	0·100	0·099	94·1	0·47
S_2	0·504	0·152	0·146	93·3	0·58	0·504	0·089	0·090	95·5	0·57

Four estimators are computed: F, full-data; CC, unstratified case-cohort; S_1, stratified with $\eta = \nu = 0{\cdot}7$; S_2, stratified with $\eta = \nu = 0{\cdot}9$. SEE, estimated standard error of the estimate. Cover., coverage of the approximate 95% confidence interval. RE, estimated efficiency relative to F.

confidence intervals have reasonable coverage rates. These results imply that the asymptotic approximations given in §§ 2 and 3 are adequate for sample size of 5000 with 500 cases. Another simulation study suggested that the asymptotic variance formulae can be used with as few as 50 cases.

The relative efficiencies given in the last column of Table 2 show that with simple random sampling a stratified estimator may be more efficient even when the whole cohort is balanced in the exposure. For example, with $\alpha = 0{\cdot}1$ and $\beta_0 = 0$, the efficiency of the unstratified estimator CC is 0·45, while the two stratified estimators S_1 and S_2 have efficiencies 0·48 and 0·59, respectively. If the subcohort had been selected by Bernoulli sampling, all the estimators would have had efficiency 0·45.

6. A REAL EXAMPLE

To illustrate the use of the stratified case-cohort design for the additive hazards regression, we now present an analysis of two related epidemiological studies, NWTSG-3 (D'Angio et al., 1989) and NWTSG-4 (Green et al., 1998). These studies were conducted by the National Wilms Tumor Study Group to study Wilms tumour, a rare renal cancer occurring in children. The most important prognostic factors for death in Wilms tumour patients are histological type and stage. Stage I–IV measures the tumour spread; histological type can be classified as favourable, FH, or unfavourable, UH. In the studies, two assessments of histological type were available, local, made in the hospital where a patient was treated, and central, made by an experienced pathologist. The local UH can be regarded as a surrogate for the true exposure, the central UH. The full cohort consisted of 4335 patients, 427 of whom had died as of the date of data listing. The median follow-up time was 5·6 years. Central UH was detected in 11·4% of the subjects. The sensitivity of local UH for central UH was 0·72 and the specificity was 0·98.

We considered a model that included five binary indicators as covariates: central UH, Stages II–IV and NWTSG-4. The last covariate was included because patients in NWTSG-4 received a better treatment compared to NWTSG-3. In the dataset, all these covariates were measured. Thus, we started by fitting the additive hazards model to the whole cohort. To emulate the case-cohort design, we drew 1000 random subcohorts from the cohort, consisting of 10% or 20% of the total sample. We did this in two ways, unstratified simple random sampling and stratified simple random sampling with strata defined by local histology status, that is favourable, unfavourable or unknown. We drew the same number of subjects from FH and UH and 10% or 20% of subjects with unknown histology. By stratification, we hoped to gain efficiency by obtaining a subcohort more balanced in central histology status.

Table 3 summarises the results. The first row gives the full-data estimates of the five parameters and their estimated standard errors. All are highly significant. The estimate for central UH was 0·0667, which means that patients with UH had on average 6·7 more deaths per 100 person-years of follow-up. The following lines show the averages of the case-cohort estimates over the 1000 simulated subcohorts and their averaged estimated standard errors. All the averaged case-cohort estimates are close to the full-data estimates. Their standard errors decrease as the subcohort size increases. When stratified sampling is used, the standard errors for central UH markedly decrease and get closer to the full-data standard error. Stratification has little or no effect on the standard errors of the other parameter estimates.

Table 3. *Results of the additive hazards regression for the NWTSG studies*

Estimate	UH (central)	Stage II	Stage III	Stage IV	NWTSG-4
			Parameter estimates		
Full-data	0·0667	0·00666	0·0152	0·0373	−0·00563
	(0·00571)	(0·00154)	(0·00196)	(0·00386)	(0·00195)
		Subcohort proportion $\alpha = 0{\cdot}1$			
Unstrat.	0·0695	0·00683	0·0154	0·0384	−0·00589
	(0·01483)	(0·00329)	(0·00387)	(0·00827)	(0·00359)
Stratif.	0·0680	0·00681	0·0156	0·0391	−0·00587
	(0·01084)	(0·00293)	(0·00354)	(0·00960)	(0·00362)
		Subcohort proportion $\alpha = 0{\cdot}2$			
Unstrat.	0·0679	0·00679	0·0155	0·0378	−0·00584
	(0·01046)	(0·00248)	(0·00296)	(0·00610)	(0·00278)
Stratif.	0·0673	0·00674	0·0155	0·0383	−0·00587
	(0·00843)	(0·00224)	(0·00275)	(0·00690)	(0·00282)

Standard error estimates for the full-data estimates are shown in parentheses. Case-cohort estimates were averaged over 1000 simulated subcohorts, with the averaged standard error estimates shown in parentheses.

7. Discussion

The proposed additive hazards regression for case-cohort studies allows the subcohort to be selected by Bernoulli sampling with arbitrary selection probabilities or by possibly stratified simple random sampling. For stratified design, simple random sampling is more efficient than Bernoulli sampling. Bernoulli sampling, however, allows more general designs. This may be desirable if, for example, the subcohort is selected ad hoc and the implicit sampling probabilities are ascertained retrospectively.

The results in §§ 4–6 show that efficiency may be improved substantially by stratifying on a surrogate exposure. The variance formulae given in Appendix 3 enable one to determine the sample size required for stratified case-cohort designs.

The proposed approach makes full use of covariate information from both the cases and controls. In particular, the covariates from all the cases are included in $\bar{Z}_H$ regardless of whether or not they belong to the subcohort. This is in the same vein as the Kalbfleisch–Lawless estimator for proportional hazards regression. There is no sensible analogue of Prentice's original estimator for additive hazards regression because both cases and controls contribute to the outside sum of the pseudo-score and they must be weighted by the same weights as in $\bar{Z}_H$. The inclusion of all cases' covariates in $\bar{Z}_H$ and in its counterpart of the Kalbfleisch–Lawless estimator makes the theoretical development more difficult because such terms are not predictable. The techniques developed in Appendix 1 do not require predictability and can also be used to establish the asymptotic properties of the Kalbfleisch–Lawless estimator.

The definition of $Y_i(.)$ given in § 2 pertains to right-censored data. In many epidemiological cohort studies, the failure time may be subject to both left-truncation and right-censoring. To accommodate left-truncation, we redefine $Y_i(t)$ as $I(X_i \geqslant t, L_i < t)$, where L_i is the time of left-truncation for the ith subject. It can be shown through slight refinements of the arguments of Appendix 1 that all the results in this paper hold for arbitrary $Y_i(.)$.

A weighted pseudolikelihood for estimating the parameters of the proportional hazards model in nested case-control studies was proposed by Samuelsen (1997). His approach could be combined with our methods to yield a nested case-control estimator for the additive hazards model. However, he did not prove asymptotic normality of the estimator and the methods given in our Appendix 1 would not suffice for a complete proof because of the complex dependence of the sampling indicators.

Recently, Borgan et al. (2000) proposed several weighted pseudo-scores for estimating the Cox model parameters in stratified case-cohort studies and discussed optimal selection of subcohort sampling probabilities. Their Estimator II shares a similar spirit to our estimator based on (8). They only gave a hint of the asymptotic properties of Estimator II, which can potentially be studied by the techniques given in our Appendix 1.

Acknowledgement

This work was supported by the U.S. National Institute of Health, the Grant Agency of the Czech Republic and the CEZ Research Project 'Mathematical methods in stochastics'. The authors are grateful to the referees for their comments and to the National Wilms Tumor Study Group for the dataset.

Appendix 1

Asymptotic approximation for the pseudo-score

We assume that regularity conditions similar to those of Andersen & Gill (1982, Theorem 4.1) hold. In particular, $\Lambda_0(\tau) < \infty$, $\text{pr}\{Y_1(\tau) = 1\} > 0$,

$$E(\sup[|Y_1(t)Z_1(t)^{\otimes 2}\{\beta_0^{\mathrm{T}} Z_1(t)\}^2|;\ 0 \leqslant t \leqslant \tau]) < \infty,$$

and $\Sigma_A(\beta_0)$ is positive definite. Then $n^{-1}\sum \rho_i Z_i(t)Y_i(t) \to \pi_1(t)$, $n^{-1}\sum \rho_i Y_i(t) \to \pi_0(t)$ and $\bar{Z}_H(t) \to e(t)$ uniformly in $t \in [0, \tau]$ in probability, where $\pi_0(t) = E\{Y_1(t)\}$ and $\pi_1(t) = E\{Z_1(t)Y_1(t)\}$. The uniformity can be shown, for example, by Corollary III.2 of Andersen & Gill (1982). Note that $\bar{Z}(t)$ also converges to $e(t)$.

We need to show that

$$n^{-\frac{1}{2}} \int_0^\tau \{e(t) - \bar{Z}_H(t)\}\, d \sum_{i=1}^n \rho_i M_i(t) = o_P(1). \tag{A1}$$

Martingale theory does not apply because ρ_i involves δ_i's and is thus not predictable. We will appeal to some results from empirical process theory.

First, let us deal with the case of independent Bernoulli sampling. Without loss of generality, assume that $Z_i(t) \geqslant 0$ for all t; otherwise, decompose each $Z_i(.)$ into its positive and negative parts. For each i, the process $\rho_i M_i(t)$ has mean zero and can be expressed as the sum of three monotone processes on $[0, \tau]$. Thus, by van der Vaart & Wellner (1996, Example 2.11.16), $B_n(t) := n^{-\frac{1}{2}} \sum \rho_i M_i(t)$ converges weakly to a tight Gaussian process $B(t)$ with continuous sample paths on $[0, \tau]$.

Since $\bar{Z}_H(t)$ is a product of two monotone processes which converge uniformly in probability to $\pi_1(t)$ and $\pi_0^{-1}(t)$, where $\pi_1(t)\pi_0^{-1}(t) = e(t)$, we can use the following approach to prove (A1). By the Skorokhod strong embedding theorem, convergence in probability and in distribution can be transformed into almost sure joint convergence on another probability space. The repeated application of integration by parts and the second Helly's theorem imply that

$$n^{-\frac{1}{2}} \int_0^\tau \bar{Z}_H(t)\, dB_n(t), \quad n^{-\frac{1}{2}} \int_0^\tau e(t)\, dB_n(t)$$

converge to the same limit $\int e(t)\, dB(t)$ almost surely on the new probability space and hence in probability on the original space. This proves (A1). The readers are referred to M. Kulich's 1997 University of Washington Ph.D. dissertation for details.

To prove (A1) and hence (9) when the subcohort is selected by possibly stratified simple random sampling, we need the following proposition.

PROPOSITION. *Let $(\xi_1, \ldots, \xi_n)$ be a random vector containing $\tilde{n}$ ones and $n-\tilde{n}$ zeros, with each permutation equally likely. Let $A_1(t), \ldots, A_n(t)$ be independent and identically distributed random processes on $[0, \tau]$ with nondecreasing sample paths, where $E\{A_1(0)\}^2 < \infty$ and $E\{A_1(\tau)\}^2 < \infty$. Then*

$$W_n := n^{-\frac{1}{2}} \sum_{i=1}^{n} \xi_i \{A_i - E(A_i)\}$$

converges weakly to a tight Gaussian process.

The finite-dimensional convergence of W_n follows from Hájek's (1960) central limit theorem while the tightness follows from Example 3.6.14 of van der Vaart & Wellner (1996). The proposition implies the uniform convergence of $n^{-1}\sum \rho_i Y_i(t)$ and $n^{-1}\sum \rho_i Z_i(t) Y_i(t)$ to $\pi_0(t)$ and $\pi_1(t)$, respectively. Consequently, $\bar{Z}_H(t) \to e(t)$ uniformly in t in probability. Moreover, the proposition can be applied within each of the K strata to show that $B_n(t) := n^{-\frac{1}{2}} \sum \rho_i M_i(t)$ converges weakly to $B^*(t)$, which is a tight Gaussian process with continuous sample paths. The rest of the proof follows the arguments used in the Bernoulli case.

APPENDIX 2

Weak convergence of the cumulative baseline hazard estimator

Simple algebraic manipulation yields

$$\hat{\Lambda}_H(t) - \Lambda_0(t) = \sum_{i=1}^{n} \rho_i \int_0^t \frac{1}{\sum_{j=1}^{n} \rho_j Y_j(s)} dM_i(s) - (\hat{\beta}_H - \beta_0)^{\mathrm{T}} h(t) - (\hat{\beta}_H - \beta_0)^{\mathrm{T}} \int_0^t \{\bar{Z}_H(s) - e(s)\}\, ds.$$

The third term is obviously $o_P(n^{-\frac{1}{2}})$ uniformly in t. By Taylor expansion,

$$\hat{\beta}_H - \beta_0 = (nD_A)^{-1} U_H(\beta_0) + o_P(n^{-\frac{1}{2}}).$$

By the arguments of Appendix 1, we have

$$n^{\frac{1}{2}} \sum_{i=1}^{n} \rho_i \int_0^t \frac{1}{\sum_{j=1}^{n} \rho_j Y_j(s)} dM_i(s) = n^{-\frac{1}{2}} \int_0^t \frac{1}{\pi_0(s)} d \sum_{i=1}^{n} \rho_i M_i(s) + o_P(1),$$

where the right-hand side converges weakly to a Gaussian process. Therefore,

$$\begin{aligned} n^{\frac{1}{2}}\{\hat{\Lambda}_H(t) - \Lambda_0(t)\} = {} & n^{-\frac{1}{2}} \int_0^t \frac{1}{\pi_0(s)} d \sum_{i=1}^{n} \rho_i M_i(s) - h^{\mathrm{T}}(t) D_A^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau} \{Z_i(u) - e(u)\}\, dM_i(u) \\ & + h^{\mathrm{T}}(t) D_A^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n} (1-\delta_i)(1-\xi_i/p_i) S_i(\beta_0) + o_P(1). \end{aligned}$$

This implies the weak convergence of the left-hand side to a zero-mean Gaussian process. The calculation of the covariance function is now straightforward.

APPENDIX 3

Limiting variance formulae

Under the conditions stated in § 4, the limiting variance of the normalised full-data pseudo-score is given by

$$\begin{aligned} \Sigma_A(\beta_0) = {} & p_Z(1-p_Z) \int_0^1 \frac{\exp\{-\Lambda_0(t) - \beta_0 t\}}{p_Z \exp(-\beta_0 t) + 1 - p_Z} d\Lambda_0(t) \\ & + \beta_0 p_Z (1-p_Z)^2 \int_0^1 \frac{\exp\{-\Lambda_0(t) - \beta_0 t\}}{\{p_Z \exp(-\beta_0 t) + 1 - p_Z\}^2} dt. \end{aligned}$$

For $\beta_0 = 0$, $\Sigma_A(0) = p_Z(1-p_Z)[1-\exp\{-\Lambda_0(1)\}]$. This is based on the results of Lin & Ying (1994) and some straightforward calculations.

Write

$$K_0 = \nu\left\{\frac{1-p(0)}{p(0)}\right\} + (1-\nu)\left\{\frac{1-p(1)}{p(1)}\right\}, \quad K_1 = (1-\eta)\left\{\frac{1-p(0)}{p(0)}\right\} + \eta\left\{\frac{1-p(1)}{p(1)}\right\},$$

$$E_1(\beta_0) = p_Z\int_0^1 \{(1-p_Z)\exp(\beta_0 t) + p_Z\}^{-1}\, d\Lambda_0(t),$$

$$E_2(\beta_0) = \Lambda_0(1) - E_1(\beta_0) + \beta_0 - \log\{p_Z\exp(-\beta_0) + 1 - p_Z\},$$

for $\beta_0 \neq 0$ and $E_1(0) = p_Z\Lambda_0(1)$, $E_2(0) = (1-p_Z)\Lambda_0(1)$. By simple algebra, we have

$$\Sigma_{HB}(\beta_0) = K_0(1-p_Z)\exp\{-\Lambda_0(1)\}E_1^2(\beta_0) + K_1 p_Z\exp\{-\Lambda_0(1) - \beta_0\}E_2^2(\beta_0).$$

Specifically, $\Sigma_{HB}(0) = p_Z(1-p_Z)\{p_Z K_0 + (1-p_Z)K_1\}\Lambda_0^2(1)\exp\{-\Lambda_0(1)\}$.

It is easy to show that $\Sigma_{HS} = \Sigma_{HB}(\beta_0) - Q(\beta_0)$, where

$$Q(\beta_0) = \frac{1-p(0)}{p(0)(1-p_V)}\{(1-\eta)p_Z\exp(-\beta_0)E_2(\beta_0) - \nu(1-p_Z)E_1(\beta_0)\}^2 + \frac{1-p(1)}{p(1)p_V}\{\eta p_Z\exp(-\beta_0)E_2(\beta_0) - (1-\nu)(1-p_Z)E_1(\beta_0)\}^2.$$

In particular,

$$Q(0) = \left\{\frac{1-p(0)}{p(0)(1-p_V)} + \frac{1-p(1)}{p(1)p_V}\right\}[p_Z(1-p_Z)(\eta+\nu-1)\Lambda_0(1)\exp\{-\Lambda_0(1)\}]^2.$$

Substituting $\nu = \eta = p_V = \frac{1}{2}$ into the expressions for Σ_{HS} and Σ_{HB}, we obtain the limiting pseudo-score variances for unstratified sampling.

REFERENCES

ANDERSEN, P. K. & GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100–20.

BORGAN, Ø., GOLDSTEIN, L., LANGHOLZ, B., POGODA, J. & SAMUELSEN, S. O. (2000). Exposure stratified case-cohort designs. *Lifetime Data Anal.* **6**. To appear.

BRESLOW, N. E. & DAY, N. E. (1987). *Statistical Models in Cancer Research,* **2**: *The Design and Analysis of Cohort Studies*. Lyon: IARC.

COX, D. R. & OAKES, D. (1984). *Analysis of Survival Data*. London: Chapman and Hall.

D'ANGIO, G. J., BRESLOW, N., BECKWITH, J. B. et al. (1989). Treatment of Wilms tumor: Results of the third National Wilms Tumor Study. *Cancer* **64**, 349–60.

GREEN, D. M., BRESLOW, N. E., BECKWITH, J. B. et al. (1998). Comparison between single-dose and divided-dose administration of dactinomycin and doxorubicin for patients with Wilms' tumor: A report from the National Wilms' Tumor Study Group. *J. Clin. Oncol.* **16**, 237–45.

HÁJEK, J. (1960). Limiting distributions in simple random sampling from a finite population. *Pub. Math. Inst. Hungar. Acad. Sci.* **5**, 361–74.

HORVITZ, D. G. & THOMPSON, D. J. (1951). A generalization of sampling without replacement from a finite universe. *J. Am. Statist. Assoc.* **47**, 663–85.

KALBFLEISCH, J. D. & LAWLESS, J. F. (1988). Likelihood analysis of multi-state models for disease incidence and mortality. *Statist. Med.* **7**, 149–60.

LIN, D. Y. & YING, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–71.

PRENTICE, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73**, 1–11.

SAMUELSEN, S. O. (1997). A pseudolikelihood approach to analysis of nested case-cohort studies. *Biometrika* **84**, 379–394.

SELF, S. G. & PRENTICE, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann. Statist.* **16**, 64–81.

VAN DER VAART, A. W. & WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.

[*Received August* 1998. *Revised September* 1999]